Chapter 7

Comparing the GLM-Models with Standard Change Detection Methods

Introduction

This chapter will continue with the example presented in the previous chapter. The intention of this chapter is to compare the change detection models developed with GLMs to standard change detection procedures. For this we will need something with which to compare. For this reason the chapter starts with a digressing to present a change detection based on "standard" methods. We will use image algebra as our “standard” change detection algorithm and the traditional error matrix as our “standard” accuracy assessment tool. For the image algebra we use the difference in normalized difference vegetation index (NDVI). There is nothing in particular or special about using the difference in NDVI except that it is a standard method. Again, our development of the NDVI change detection is simply to give us something against which we can compare the GLMs developed in the previous chapter. The main focus of this chapter is not so much that the logistic models are more accurate than the NDVI-based model but *how to compare* the logistic models with a standard change detection algorithm.

The chapter starts with a description of the change detection method based on NDVI and then presents a traditional accuracy assessment for this change detection using the error matrix. We go on to discuss the difficulties of using the error matrix to compare models. The SAS™ output from the LOGISTIC procedure can be used to produce a more informative accuracy assessment using what we refer to as "accuracy assessment curves". To demonstrate the use of the accuracy assessment curves, we present these curves for the

NDVI-based and the Logistic models. Using these curves we find the logistic models are more accurate. In this comparison we see that the concept of the accuracy assessment curves can be helpful for comparing two change detection algorithms as well as for assessing the accuracy of a particular change detection algorithm.

Using NDVI as an Example Standard Change Detection

The NDVI is seen as an indicator of healthy vegetation (Jensen, 1996, pp. 181 - 182; Chavez and MacKinnon, 1994). Differences in NDVI have been used to detect changes due to differences in vegetation (Townshend *et al*., 1992; Choung, 1992). Equation 6.3 in the previous chapter gives the formula of the difference in NDVI. Using the difference in NDVI is a particular example of the "image algebra" change algorithm discussed in chapter 4. For each date an NDVI image is constructed. Then we subtract the 1988 NDVI image from the 1994 NDVI image to produce an NDVI-difference image. By the nature of the index, when we subtract the 1988 NDVI image from the 1994 NDVI image we would expect planting and growth of vegetation to result in positive values and cutting or clearing of vegetation to result in negative values. The histogram of the NDVI-differences image for the coastal area is shown in figure 7.1 and the histogram of the NDVI-differences image for the Raleigh area is shown in figure 7.2. The summary statistics for these two difference images are given in table 7.1.

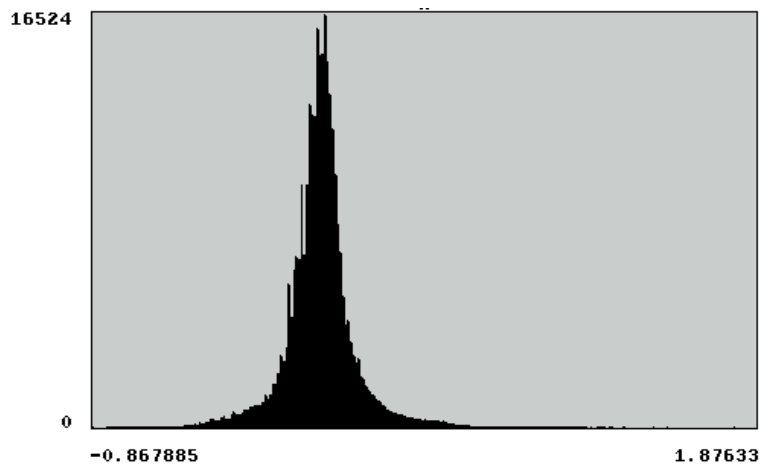

Figure 7.1: Histogram for the NDVI difference image, coastal area

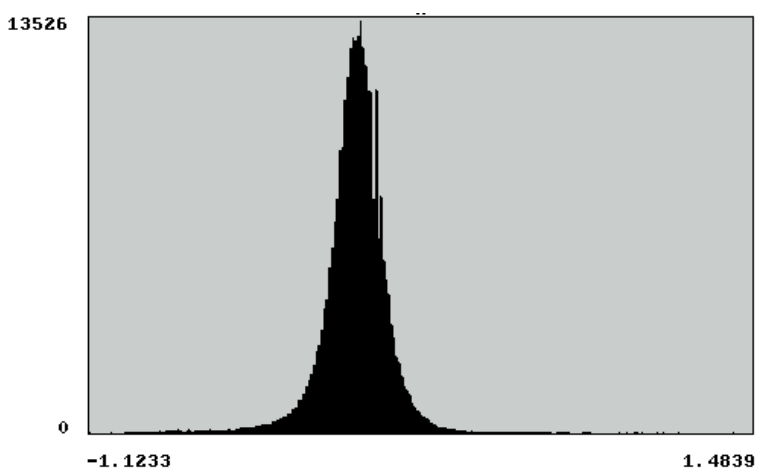

Figure 7.2: Histogram for the NDVI difference image, Raleigh area

Table 7.1: Summary statistics for the NDVI difference images

Area	Average	Median	Standard Deviation	Minimum	Maximum
Coastal scenes	0.071	0.06472	0.142	-0.8678	1.8763
Raleigh scenes	-0.081	-0.074312	0.129	-1.1233	1.48391

A typical approach to the image algebra algorithm is to examine the histogram of the difference image and set a change threshold based on units of standard deviations from the mean (Jensen, 1996, p. 269). With the image algebra change detection procedure the model of change is a simple step function where the "step" is located at the threshold value. This step function for one standard deviation is shown in figure 7.3.

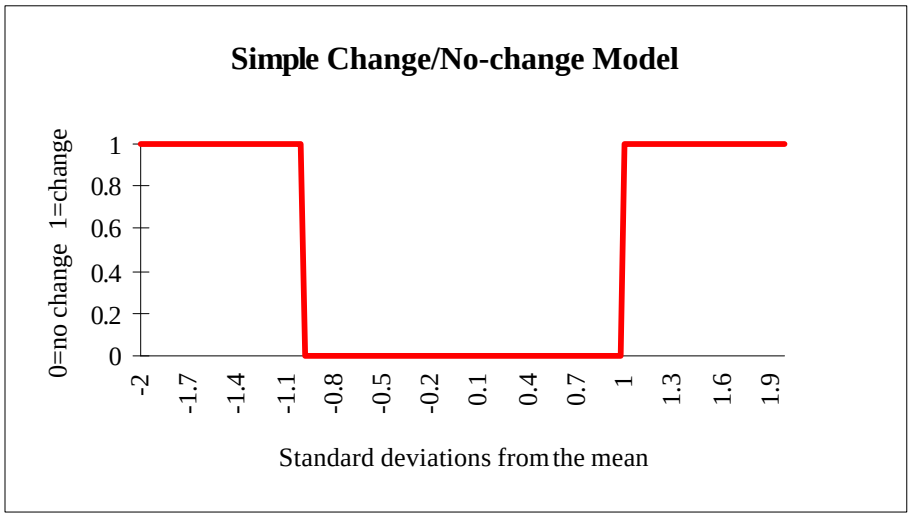

Figure 7.3: Image algebra change detection step function model

Using the Error Matrix as a Standard Accuracy Assessment Tool

We now move to an accuracy assessment of the NDVI-based change detection algorithm. For this accuracy assessment we will use the standard method of producing an error matrix. There is some additional discussion given to help tie this section to the following section on accuracy assessment curves.

Since all image algebra change detection algorithms require a threshold selection, and because the location of the threshold can effect the accuracy (Fung and LeDrew, 1988), we will consider five change thresholds. These will be set at ±0.5, ±0.75, ±1.0, ±1.5 and ±2 standard deviations from the mean. For each point in our sample (described in Chapter 6) we can check whether the point has changed or not (based on our reference data) and

compare the NDVI values (based on the image data) for that point. From this we can determine how each point would be classified at the different threshold levels of NDVI-differences. We use this to produce an error matrix for each threshold level. These error matrices are two-by-two tables where the rows represent the reference data and the columns represent classification based on differences in NDVI. The error matrix is a typical method used to assess classification accuracy (Congalton, 1991). From each error matrix, we will extract several statistics:

- Overall accuracy, which is the sum of the diagonal elements divided by the total number in the sample
- The percentage correct for a given row divided by the total for that row, referred to in the remote sensing literature as "producers accuracy" (Lillesand and Kiefer, 1994, p. 613). In statistical literature terms the producer's accuracy for the *no-change* sample points can be through of as the model's "specificity" and the producer's accuracy for the *change* sample points can be considered the model's "sensitivity".
- The percentage correct for a given column divided by the total for that column, referred to in the remote sensing literature as "users accuracy" (Lillesand and Kiefer, 1994, pp. 613).
- KHAT coefficient, which is an index that relays the classification accuracy after adjustment for chance agreement. KHAT is the estimate of the Kappa coefficient (Cohen, 1960; Congalton *et al.*, 1983).

The KHAT equation is computed as (Congalton, 1991):

$$\boldsymbol{KHAT} = \frac{\boldsymbol{N}\sum_{i=1}^{2} \boldsymbol{x}_{ii} - \sum_{i=1}^{2} (\boldsymbol{x}_{i+} \times \boldsymbol{x}_{+i})}{\boldsymbol{N}^{2} - \sum_{i=1}^{2} (\boldsymbol{x}_{i+} \times \boldsymbol{x}_{+i})}$$

where

$\boldsymbol{x_{ii}}$ = the diagonal element in row *i* column *I*
$\boldsymbol{x_{i+}}$ = the tth row total
$\boldsymbol{x_{+i}}$ = the tth column total
$\boldsymbol{N}$ = the total number of observation included in the matrix

Tables 7.2 and 7.3 present the error matrices for the five different threshold levels for the two study areas.

Table 7.2: Error matrices for the coastal area NDVI-based change detection

Model bases on threshold at 0.5 standard deviations

		no-change	change	Totals	Producer's Accuracy	
Observed	no-change	74	117	191	0.39	
	change	11	27	38	0.71	
	Totals	85	144	229		
	User's Accuracy	0.87	0.19		Overall Accuracy:	0.44
					Kappa Coefficient:	0.05

Model bases on threshold at 0.75 standard deviations

		no-change	change	Totals	Producer's Accuracy	
Observed	no-change	135	56	191	0.71	
	change	15	23	38	0.61	
	Totals	150	79	229		
	User's Accuracy	0.90	0.29		Overall Accuracy:	0.69
					Kappa Coefficient:	0.22

Model bases on threshold at 1 standard deviation

		no-change	change	Totals	Producer's Accuracy	
Observed	no-change	167	24	191	0.87	
	change	19	19	38	0.50	
	Totals	186	43	229		
	User's Accuracy	0.90	0.44		Overall Accuracy:	0.81
					Kappa Coefficient:	0.36

Model bases on threshold at 1.5 standard deviations

		no-change	change	Totals	Producer's Accuracy	
Observed	no-change	185	6	191	0.97	
	change	25	13	38	0.34	
	Totals	210	19	229		
	User's Accuracy	0.88	0.68		Overall Accuracy:	0.86
					Kappa Coefficient:	0.39

Model bases on threshold at 2.0 standard deviations

		no-change	change	Totals	Producer's Accuracy	
Observed	no-change	190	1	191	0.99	
	change	28	10	38	0.26	
	Totals	218	11	229		
	User's Accuracy	0.87	0.91		Overall Accuracy:	0.87
					Kappa Coefficient:	0.36

Table 7.3: Error matrices for the Raleigh area NDVI-based change detection

Model bases on threshold at 0.5 standard deviations

		no-change	change	Totals	Producer's Accuracy	
Observed	no-change	101	137	238	0.42	
	change	7	15	22	0.68	
	Totals	108	152	260		
	User's Accuracy	0.94	0.10		Overall Accuracy:	0.45
					Kappa Coefficient:	0.03

Model bases on threshold at 0.75 standard deviations

		no-change	change	Totals	Producer's Accuracy	
Observed	no-change	171	67	238	0.72	
	change	10	12	22	0.55	
	Totals	181	79	260		
	User's Accuracy	0.94	0.15		Overall Accuracy:	0.70
					Kappa Coefficient:	0.12

Model bases on threshold at 1 standard deviation

		no-change	change	Totals	Producer's Accuracy	
Observed	no-change	214	24	238	0.90	
	change	12	10	22	0.45	
	Totals	226	34	260		
	User's Accuracy	0.95	0.29		Overall Accuracy:	0.86
					Kappa Coefficient:	0.28

Model bases on threshold at 1.5 standard deviations

		no-change	change	Totals	Producer's Accuracy	
Observed	no-change	231	7	238	0.97	
	change	16	6	22	0.27	
	Totals	247	13	260		
	User's Accuracy	0.94	0.46		Overall Accuracy:	0.91
					Kappa Coefficient:	0.30

Model bases on threshold at 2.0 standard deviations

		no-change	change	Totals	Producer's Accuracy	
Observed	no-change	236	2	238	0.99	
	change	18	4	22	0.18	
	Totals	254	6	260		
	User's Accuracy	0.93	0.67		Overall Accuracy:	0.92
					Kappa Coefficient:	0.26

The error matrices presented above show how the accuracy figures will change depending on the threshold level. Fung and LeDrew (1988) also found this effect in their work on the determination of optimal threshold level for change detection. The main idea in their work and what we see here is that the location of the threshold level will affect the accuracy assessment figures.

For convenience in the following discussion we will refer to a "liberal threshold" as one which classifies more areas as changed and a "conservative threshold" as one which classifies fewer areas as changed. For example, setting the change threshold at ±0.5 standard deviations is a more liberal change threshold than the others. A change threshold at ±2 standard deviations is the most conservative change threshold considered in Tables 7.2 and 7.3.

Consider the two types of error from a simple "change/no-change" classification (i.e the off-diagonal elements of the error matrix). An error is either an area that has changed and is misclassified as "not changed" or an area that has not changed but is classified as "changed". The more liberal the change threshold the fewer misclassifications of areas that have changed but the more misclassifications of areas which have not. For a more conservative change threshold, more points that have changed are erroneously classified and more unchanged sample points are correctly classified. To make this point more clear, consider the extremes. For our NDVI-difference image, set an extremely liberal change threshold of only ±0.1 standard deviations from the mean. With such a threshold, nearly the entire image would be classified as "changed" and you would not have many errors with respect to misclassifying areas that have changed. However, there would be a large amount of misclassifications in areas that have not changed. Now consider the other extreme. Set the change threshold at ± five standard deviations from the mean. This threshold would classify almost all of the points as "not changed". So, for this classification there would be few errors with respect to points which have not changed but nearly all of those points which have changed would be misclassified as "unchanged".

In selecting the threshold level one needs to be careful not to base the decision on just one of the accuracy figures. If overall accuracy is used to determine the best threshold level, then in our example both the coastal and Raleigh areas change threshold would be set at ± two standard deviations. However, this is the worst setting with respect to the user's accuracy for change areas. That means the reference sample points that have changed have the lowest chance of being classified correctly when the threshold is set at 2 standard deviations. For our data, by selecting a threshold that maximized the overall accuracy, we inadvertently minimize the chance of classifying correctly those areas that have changed. Fung and LeDrew (1988) recommend using the KHAT coefficient for determining the optimal threshold because it uses all cells of the error matrix. However, instead of using one element from the accuracy assessment to base the selection of the threshold level, it is better to understand the *relationship* between the threshold level and the different accuracy figures. Understanding this relationship provides a deeper understanding of the change detection model as well as the change detection product itself.

Before we go on to compare the NDVI-based classification to the logistic change model we note that we have not actually used the logistic models to classify the image into change/no-change areas. The logistic model produces a surface of the "probability of change". In order to compare the logistic change model to the NDVI-based model we need to use the logistic model to classify the image. This can be done by selecting a probability value to use as a cut off or threshold level. Here again we have the same issue of selecting a threshold value and then interpreting the result of the subsequent accuracy assessment based on different threshold levels. This brings us back to the idea of understanding the relationship between the threshold value and the accuracy figures. In order to compare the two models we need to compare the relationships between the threshold levels for each model and the accuracy figures for the model.

Accuracy Assessment Curves

These curves are a graph that shows the relationship between the threshold level used to

classify the changed areas and the different accuracy assessment figures typically contained in the error matrix. Each element from the typical accuracy assessment (for example, the overall accuracy) will have its own line within the accuracy assessment curves. This is because each element from the accuracy assessment has its own relationship with the threshold level. By plotting the curves for each element from the accuracy assessment together on one graph we can see the relationship between the different accuracy assessment elements with respect to the threshold level and each other. We will plot lines for the overall accuracy, producer’s accuracy, user's accuracy and KHAT. In comparing two models, the model with higher accuracy assessment curves is a more accurate model.

Figure 7.4 is a diagram to help with understanding the lines that are presented on the accuracy assessment curves. The figure presents a color code to match the lines on the accuracy assessment curves with the corresponding figures as they typically appear in an error matrix. A modified version of this graphic will appear on each of the accuracy assessment curves.

		Classified Data				
		no-change	change			
Reference Data	no-change	**C**	**E**	**C / Row Total**	**% Classified Correctly from "Unchanged" sample point**	**= producer's accuracy for no change**
	change	**E**	**C**	**C / Row Total**	**% Classified Correctly from "Changed" sample points**	**= producer's accuracy for change**
		C / Column Total	**C / Column Total**			
		% Correct From points Classified as "Unchanged"	**% Correct from points Classified as "Changed"**		**Overall Accuracy**	
		= user's accuracy for no change	**= user's accuracy for change**		**Kappa Coefficient**	

Figure 7.4: Diagram to accompany accuracy assessment curves

Before we present the accuracy assessment curves, we need to explain how the accuracy assessment figures for the logistic models are derived. For the logistic models, the reference data were used to estimate the model.

> When you use the same data to test the predictive accuracy of your model that you use to fit the model, it biases the results…(A) way to avoid this bias... is to fit a model that omits each observation one at a time and then classify each observation as an event or nonevent based on the model that omits the observation being classified. The method…is called jackknifing. (SAS, 1995, p. 36)

The SAS™ LOGISTIC procedure produce the unbiased jackknife estimates for the classification accuracies. By specifying the CTABLE model option in the LOGISTIC procedure, SAS™ will output a sequence of classification accuracy figures.

> The LOGISTIC procedure produces a classification table that contains several measures of predictive accuracy for each probability cutpoint. That is, the model classifies an observation as an event (i.e. change) if its estimated probability is greater than or equal to a given probability cutpoint. Otherwise, the observation is classified as a nonevent (i.e. no change). As the probability cutpoints increase in value, the more likely that an observation is classified as a nonevent (i.e. no change). The classification table reports how well these classifications match the observed event or non-event status of each observation. (SAS, 1995, p. 45)

We will use the tables produced in the SAS™ output to produce accuracy assessment curves. The alternative is to collect another set of reference data. Since the reference data collection was the most labor and time int3ensive component of the change detection procedure, we will use the unbiased estimate produced by the SAS™ system.

Using the Accuracy Assessment Curves

Tables 7.2 and 7.3 show how the accuracy assessment figures for the NDVI-based algorithm change as a function of where the threshold is set. This is consistent with the result of Fung and LeDrew, 1988; Choung, 1992). In order to get a better understanding of the relationship between the threshold level and the accuracy assessment figures we can plot the accuracy values against the threshold levels. Fung and LeDrew (1988) present an accuracy assessment curve in their evaluation of the effect of threshold levels on change detection accuracy. They recommend using the KHAT coefficient for determining the optimal threshold because it uses all cells of the error matrix. However, as stated above, it is more appropriate to consider the *relationship* shown on the accuracy assessment curves. For example, it may be that the most important aspect of the final change detection product is to accurately locate areas that have changed. This would imply choosing a threshold that results in a high producer's accuracy. Other accuracy figures would no doubt be important and so the other lines on within the accuracy assessment

curves would need to be considered. So, instead of using one term to select the "optimal" threshold level for change a change detection algorithm, it is more appropriate to consider the purpose of the change detection. The accuracy assessment curves can be used to select threshold based on the ultimate goal of the change detection.

In figures 7.5 through 8 we present the accuracy assessment curves for the logistics regression and NDVI based change detection models. For the logistic regression model the threshold would be based on different levels of probability of change and so this is the X-axis for the logistic regression accuracy assessment curves. For the NDVI-based change detection, the threshold is in unit of ± standard deviations from the mean. So, for the NDVI models, the X-axis is in units of standard deviations.

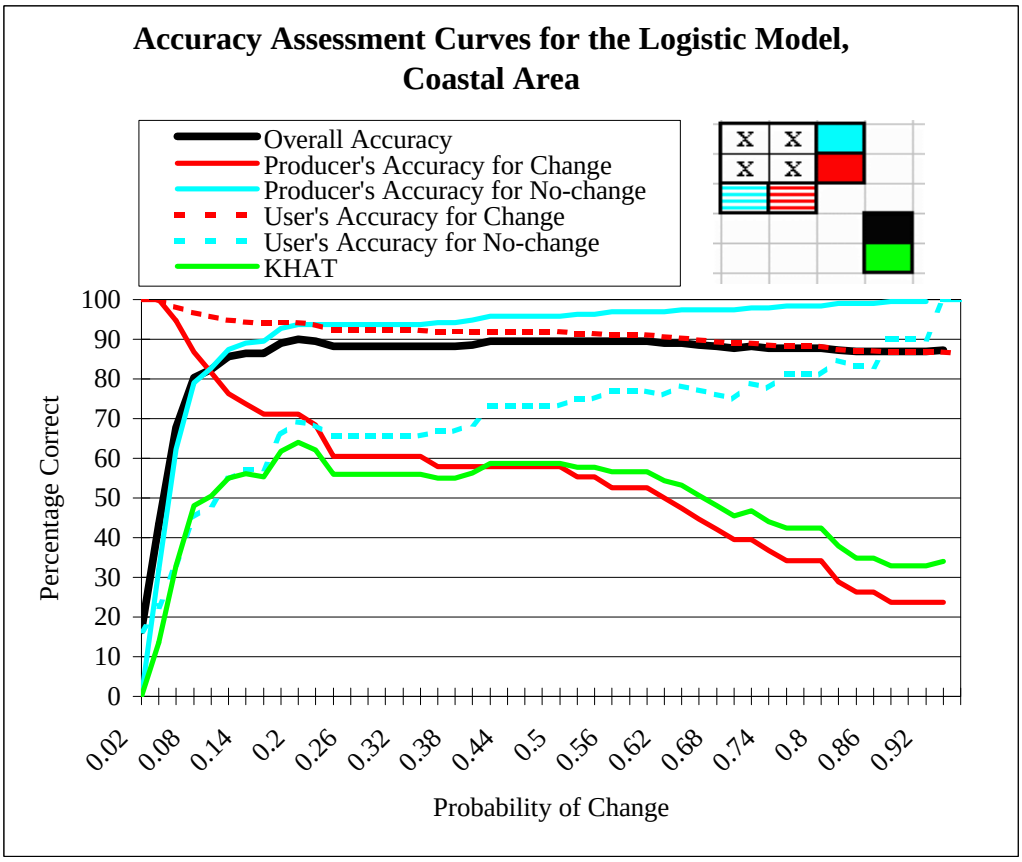

Figure 7.5: Accuracy assessment curves for the logistic regression model for the coastal area

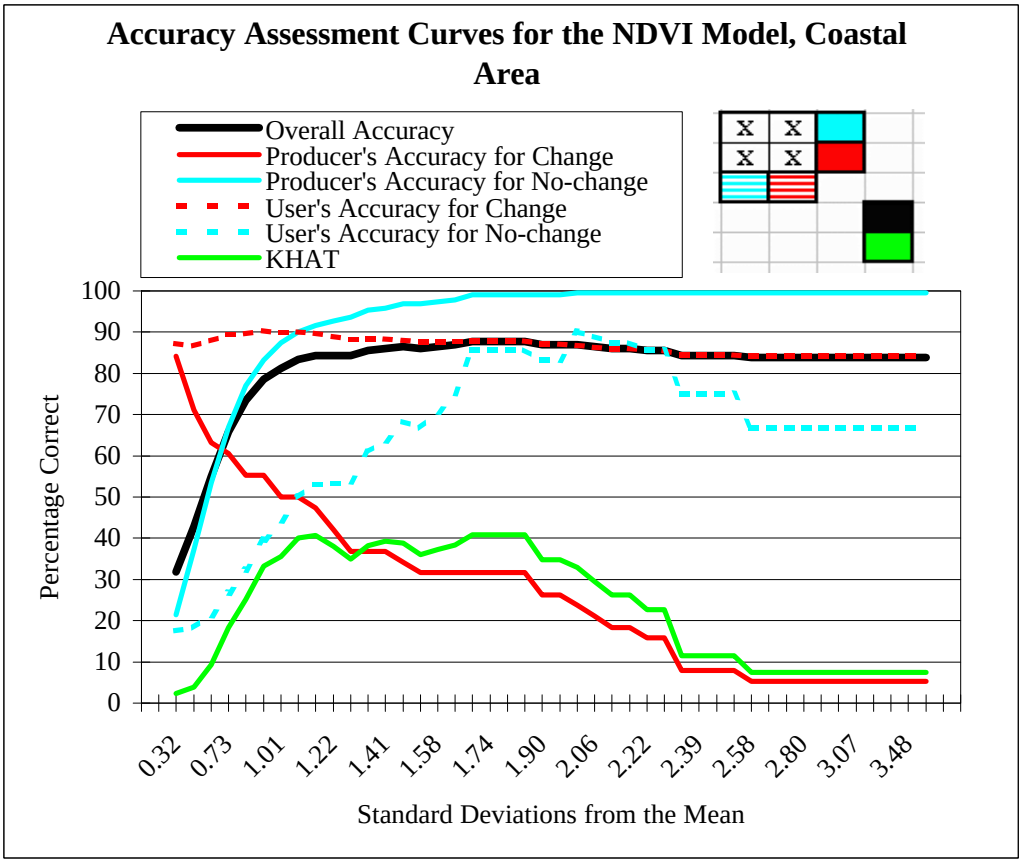

Figure 7.6: Accuracy Assessment curves for the NDVI-based model for the coastal area

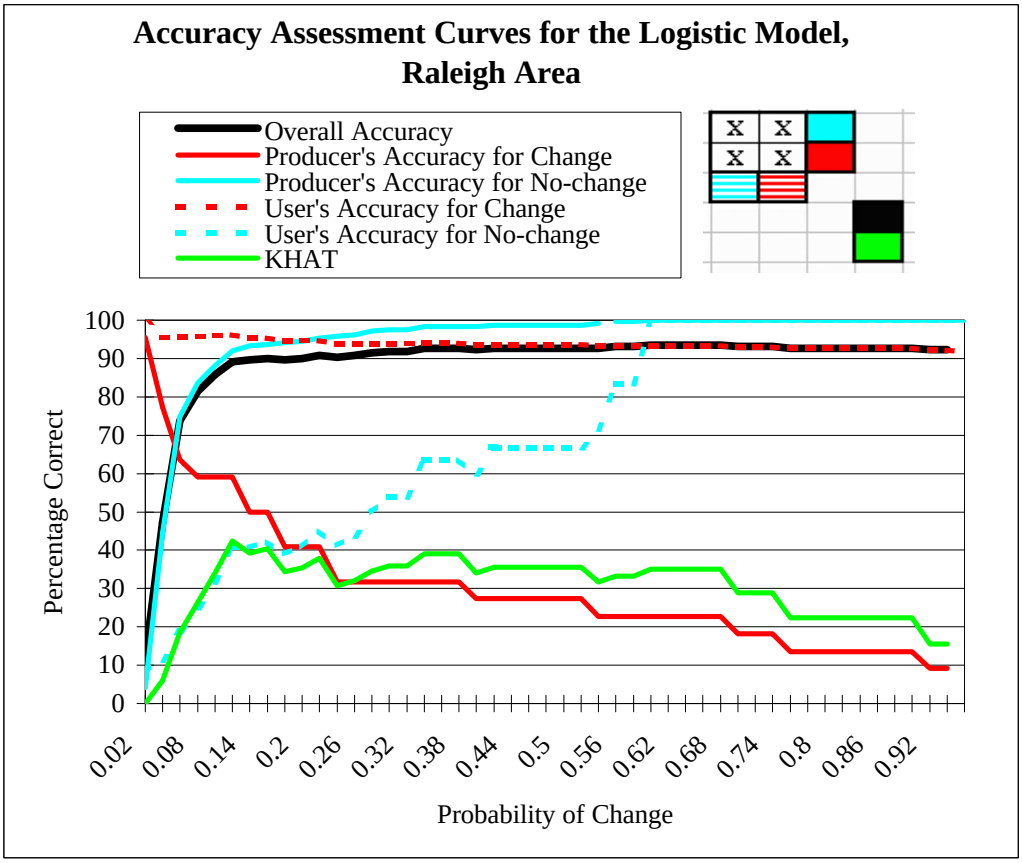

Figure 7.7: Accuracy assessment curves for the logistic regression model for the Raleigh area

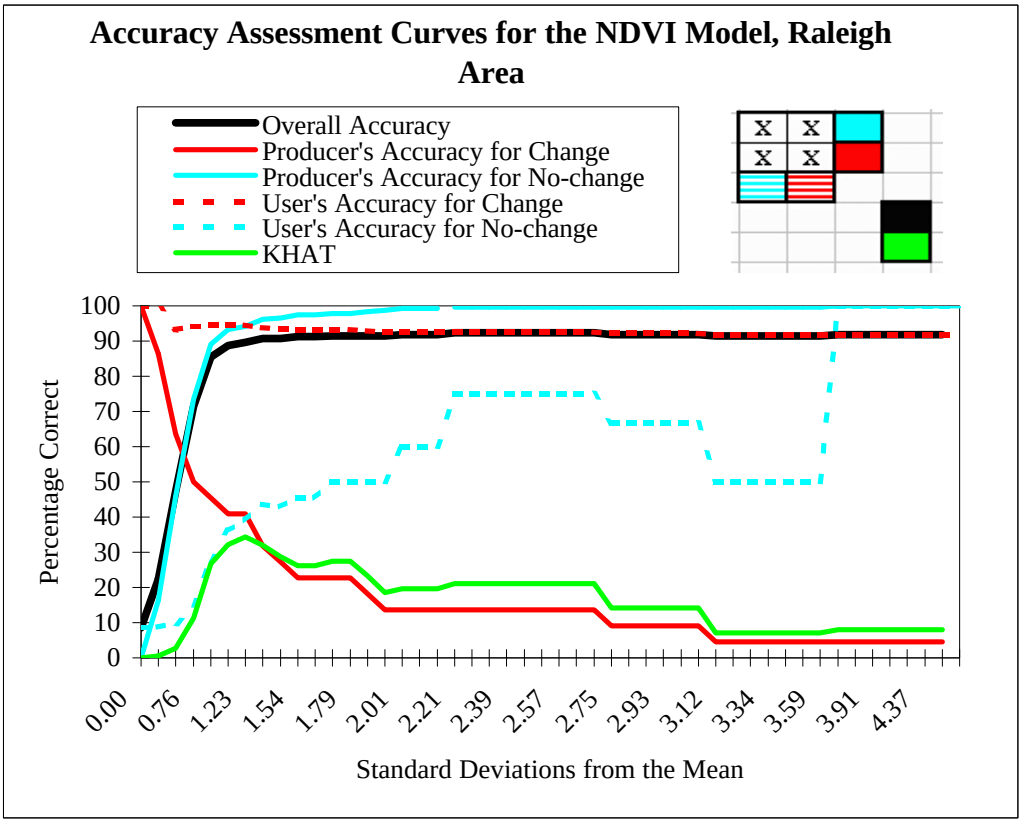

Figure 7.8: Accuracy Assessment curves for the NDVI-based model for the Raleigh area

These figures present several points of interest. First is that the lines are higher for the logistic regression models. This implies that the logistic regression models are superior to the NDVI models. That is, going across all possible thresholds, for each study area the accuracy figures for the logistic regression model are higher than the NDVI model accuracy figures. For any given producer's accuracy for change the logistic models will have a higher producer's accuracy for no-change. The second point is that the curves help relay the relationship between the threshold level and the prediction accuracy that results from that threshold. We can clearly see the inverse relationship between the producer's accuracy for change and the producer's accuracy for no-change. The curves show how much you need to compromise between the two. The third point is that these plots provide insight into the threshold selection. As alluded to earlier, the threshold selection

should be based on the accuracy assessment curves and the ultimate goal of the change detection study. If the goal is to produce an estimate of the amount of both change and no-change areas, then maximizing the KHAT coefficient, because it uses the values from the entire error matrix and adjusts for change agreement, may be the best criterion for threshold selection. If the goal is to find any and all areas that have changed in order to direct further investigation, a threshold level with a high producer's accuracy for change -- subject to some constraints on the other accuracy assessment figures -- may be the best selection criterion. The main idea is that the accuracy assessment curves should be considered in light of the objective(s) of the change detection and a threshold should be based on the objective(s) and the accuracy assessment curves. However, with the GLMs there is no need to select a threshold. Instead of producing a change/no-change classified map, we can use GLMs to produce a map of the probability of change. This idea will be addressed in the next chapter.

Conclusions from Comparing GLMs to Traditional Methods.

By comparing the GLM based change detection to a standard change detection we have found further justification for using the GLM for change detection. The LOGISTIC procedure in SAS™ can produce unbiased estimates for accuracy assessment. It does so for different threshold level used to classify change areas. The accuracy assessment figures can be plotted against the threshold levels to produce accuracy assessment curves, which show the effect of threshold level on the change detection accuracy. By using these plots we found that the accuracy for the logistic regression models developed in the previous chapter were more accurate than using the differences in NDVI. We also can use these plots to realize the significance of threshold placement on the different accuracy assessment figures. The next chapter will explain how the accuracy assessment curves can be used in conjunction with the probability of change map to produce a more informative change detection product.